

Rate-Distortion Bounds for an ε -Insensitive Distortion Measure

Kazuho Watanabe

Graduate School of Information Science

Nara Institute of Science and Technology, Nara, 630-0192, Japan

Email: wkazuho@is.naist.jp

Abstract—Direct evaluation of the rate-distortion function has rarely been achieved when it is strictly greater than its Shannon lower bound. In this paper, we consider the rate-distortion function for the distortion measure defined by an ε -insensitive loss function. We first present the Shannon lower bound applicable to any source distribution with finite differential entropy. Then, focusing on the Laplacian and Gaussian sources, we prove that the rate-distortion functions of these sources are strictly greater than their Shannon lower bounds and obtain analytically evaluable upper bounds for the rate-distortion functions. Small distortion limit and numerical evaluation of the bounds suggest that the Shannon lower bound provides a good approximation to the rate-distortion function for the ε -insensitive distortion measure.

I. INTRODUCTION

In source coding, the rate-distortion function $R(D)$ of a source shows the minimum information rate required to reconstruct the source outputs with average distortion not exceeding D . Rate-distortion functions have been explicitly evaluated for various sources and distortion measures. The Shannon lower bound (SLB) $R_L(D)$ plays an important role in the explicit evaluation of rate-distortion functions of difference distortion measures. A common approach is to derive $R_L(D)$ and examine the condition for $R(D)$ to coincide with $R_L(D)$ [4], [1]. There have been, however, only several results when $R(D) > R_L(D)$ for all D . In this case, direct explicit evaluation of $R(D)$ has been achieved only in limited cases such as discrete memoryless finite-alphabet sources [1] and a class of sources under an absolute-magnitude distortion measure [9], [12], [11]. There have also been indirect approaches. Rose proposed a deterministic annealing algorithm to generate $R(D)$ based on the fact that under the squared distortion measure, the optimal reconstruction is purely discrete when $R(D) > R_L(D)$. Buzo *et al.* obtained upper and lower bounds for $R(D)$ under the Itakura-Saito distortion measure [2].

In this paper, we focus on the ε -insensitive loss function as a distortion measure, which was introduced to support vector machines for regression [10]. We obtain the SLB for this difference distortion measure, which is analytically evaluable for arbitrary sources with finite differential entropy. Then, we examine the condition for the rate-distortion function to coincide with the SLB. Taking the Laplacian and Gaussian sources as specific examples, we prove that the rate-distortion functions of these sources lie strictly above their SLBs for all D when $\varepsilon > 0$ and derive analytically evaluable upper

bounds for the rate-distortion functions. Investigation of small distortion limit of these upper bounds shows that the SLB has the accuracy of $O(\varepsilon^2)$ as $D \rightarrow 0$ in both sources. Numerical evaluation of the lower and upper bounds demonstrates that the SLB gives a good approximation to $R(D)$ for small distortion while the trivial upper bound provided by the rate-distortion function for $\varepsilon = 0$ suggests that the SLB also gives a reasonable approximation for large distortion as well.

II. RATE-DISTORTION FUNCTION FOR THE ε -INSENSITIVE DISTORTION MEASURE

A. Rate-Distortion Function

Let X and Y be random variables on \mathbf{R} and $d(x, y)$ be the non-negative distortion measure between x and y . The rate-distortion function $R(D)$ of the source $X \sim p(x)$ with respect to the distortion d is defined by

$$R(D) = \inf_{q(y|x): E[d(x, y)] \leq D} I(q), \quad (1)$$

where

$$I(q) = \int \int q(y|x)p(x) \log \frac{q(y|x)}{\int q(y|x)p(x)dx} dx dy$$

is the mutual information and E denotes the expectation with respect to $q(y|x)p(x)$. $R(D)$ shows the minimum achievable rate for the i.i.d. source with the density $p(x)$ under the given distortion measure d [1], [4].

The above minimization problem can be reformulated as a problem of minimization over the reproduction density $q(y)$ [1], [6],

$$\inf_{q(y)} \left[- \int p(x) \log \int \exp(sd(x, y))q(y)dy dx \right], \quad (2)$$

where $s \leq 0$ is a parameter. Then, if there exists $q_s(y)$ that achieves the infimum in Eq. (2), $R(D)$ is parametrically given by

$$\begin{aligned} R(D_s) &= - \int p(x) \log \int \exp(sd(x, y))q_s(y)dy dx + sD_s, \\ D_s &= \int \int p(x)q_s(y|x)d(x, y)dx dy, \end{aligned} \quad (3)$$

where the optimal conditional density of reconstruction, $q_s(y|x)$ is defined by

$$q_s(y|x) = \frac{q_s(y) \exp(sd(x, y))}{\int q_s(y) \exp(sd(x, y))dy}. \quad (4)$$

In Eq. (3), $R(D)$ is parameterized by $s \leq 0$, which corresponds to the slope of the tangent of $R(D)$ at $(D_s, R(D_s))$ and hence is referred to as the slope parameter [1].

From the properties of the rate-distortion function $R(D)$, we know that $R(D) > 0$ for $0 < D < D_{\max}$, where

$$D_{\max} = \inf_y \int p(x) d(x, y) dx, \quad (5)$$

and $R(D) = 0$ for $D \geq D_{\max}$ [1, p. 90].

B. ε -Insensitive Loss Function

In this paper, we focus on the following difference distortion measure defined by the ε -insensitive loss function ρ_ε (Fig. 1),

$$d(x, y) = \rho_\varepsilon(x - y), \quad (6)$$

where

$$\rho_\varepsilon(z) = \begin{cases} |z| - \varepsilon, & (|z| \geq \varepsilon), \\ 0, & (|z| < \varepsilon). \end{cases}$$

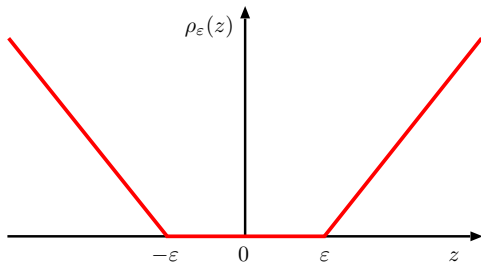


Fig. 1. The ε -insensitive loss function $\rho_\varepsilon(z)$.

This loss function with $\varepsilon > 0$ was introduced to support vector regression in order to provide a sparsity inducing mechanism [10], [8]. We denote the rate-distortion function for this distortion measure by $R^{(\varepsilon)}(D)$ and the maximum distortion D_{\max} in Eq. (5) by $D_{\max}^{(\varepsilon)}$.

C. Shannon Lower Bound

Generally for difference distortion measures, Shannon obtained a lower bound to $R(D)$, which is referred to as the Shannon lower bound (SLB) [1, p. 92]. For the ε -insensitive distortion measure, it is parametrically expressed as

$$R^{(\varepsilon)}(D_s) \geq R_L^{(\varepsilon)}(D_s) = h(p) - h(g_s), \quad (7)$$

$$D_s = \int \rho_\varepsilon(x) g_s(x) dx, \quad (8)$$

where $h(p) = -\int p(x) \log p(x) dx$ is the differential entropy of the probability density p and g_s is the probability density function defined by¹

$$g_s(x) = \frac{e^{s\rho_\varepsilon(x)}}{\int e^{s\rho_\varepsilon(z)} dz}. \quad (9)$$

We explicitly evaluate $h(g_s)$ to obtain the SLB. The density g_s is explicitly given by

$$g_s(x) = \begin{cases} \frac{1}{C_s} e^{-s(x+\varepsilon)}, & (x \leq -\varepsilon), \\ \frac{1}{C_s}, & (|x| < \varepsilon), \\ \frac{1}{C_s} e^{s(x-\varepsilon)}, & (x \geq \varepsilon), \end{cases} \quad (10)$$

¹ We omit the dependency on ε in notations unless we put $\varepsilon = 0$.

where

$$C_s = 2 \frac{1 + |s|\varepsilon}{|s|}. \quad (11)$$

Its differential entropy is evaluated as,

$$h(g_s) = \log \left(2 \frac{1 + |s|\varepsilon}{|s|} \right) + \frac{1}{1 + |s|\varepsilon}. \quad (12)$$

The slope parameter s is related to the average distortion D_s by Eq. (8), which is rewritten as,

$$D_s = \frac{2}{s^2 C_s} = \frac{1}{(1 + \varepsilon|s|)|s|}. \quad (13)$$

Solving this for $|s|$ yields

$$|s| = \frac{-D_s + \sqrt{D_s^2 + 4D_s\varepsilon}}{2D_s\varepsilon}.$$

Putting this back into Eq. (12), we have

$$h(g_s) = \log(2\varepsilon + D_s + \sqrt{D_s^2 + 4D_s\varepsilon}) - \frac{D_s - \sqrt{D_s^2 + 4D_s\varepsilon}}{2\varepsilon}.$$

Thus, from Eq. (7), we obtain the following theorem.

Theorem 1 *The rate-distortion function for the ε -insensitive distortion measure in Eq. (6) satisfies $R^{(\varepsilon)}(D) \geq R_L^{(\varepsilon)}(D)$ for all D , where*

$$R_L^{(\varepsilon)}(D) = h(p) - \log(2\varepsilon) - \log \left(1 + \tilde{D} + \sqrt{\tilde{D}^2 + 2\tilde{D}} \right) + \tilde{D} - \sqrt{\tilde{D}^2 + 2\tilde{D}},$$

$\tilde{D} = \frac{D}{2\varepsilon}$ and $h(p)$ is the differential entropy of the source density.

D. Condition for $R^{(\varepsilon)}(D) = R_L^{(\varepsilon)}(D)$

For any negative value of the slope parameter s , the lower bound $R_L^{(\varepsilon)}(D_s)$ coincides with $R^{(\varepsilon)}(D_s)$ if and only if the condition

$$p(x) = \int q(y) g_s(x - y) dy, \quad (14)$$

holds for all x and a valid density function $q(y)$ [1, p. 94].

The condition in Eq. (14) is equivalent to

$$P(\omega) = Q(\omega) G_s(\omega), \quad (15)$$

where P , Q and G_s are the Fourier transforms (characteristic functions) of p , q and g_s respectively.

The Fourier transform of g_s in Eq. (10) is specifically given by

$$\begin{aligned} G_s(\omega) &= \int \frac{e^{s\rho_\varepsilon(x)}}{C_s} e^{-i\omega x} dx \\ &= \frac{s^2}{s^2 + \omega^2} \cdot \frac{\varepsilon|s| \frac{\sin(\omega\varepsilon)}{\omega\varepsilon} + \cos(\omega\varepsilon)}{1 + |s|\varepsilon} \\ &\equiv L_{|s|}(\omega) \cdot M_{|s|}^{(\varepsilon)}(\omega). \end{aligned} \quad (16)$$

Here, the first factor, defined as $L_{|s|}(\omega) = \frac{s^2}{s^2 + \omega^2}$, is the characteristic function of the Laplace distribution with parameter $|s|$ whose density function is $l_{|s|}(x) = \frac{|s|}{2} e^{-s|x|}$. The second

factor, $M_{|s|}^{(\varepsilon)}(\omega) = \frac{\varepsilon|s| \frac{\sin(\omega\varepsilon)}{\omega\varepsilon} + \cos(\omega\varepsilon)}{1 + |s|\varepsilon}$, is the characteristic function of the mixture of the delta distributions (on $-\varepsilon$ and ε) and the uniform distribution on $[-\varepsilon, \varepsilon]$ mixed with the proportion $1 : \varepsilon|s|$. More specifically, the density function of this mixture is expressed as

$$m_{|s|}^{(\varepsilon)}(x) = \frac{1}{1 + \varepsilon|s|} \frac{\delta(x - \varepsilon) + \delta(x + \varepsilon)}{2} + \frac{\varepsilon|s|}{1 + \varepsilon|s|} u_{[-\varepsilon, \varepsilon]}(x),$$

where δ is the Dirac delta function and $u_{[-\varepsilon, \varepsilon]}$ is the density function of the uniform distribution on $[-\varepsilon, \varepsilon]$. Hence, Eq. (16) means that the density g_s is given by the convolution $l_{|s|} * m_{|s|}^{(\varepsilon)}$ of $l_{|s|}$ and $m_{|s|}^{(\varepsilon)}$. Summarizing Eqs. (15) and (16), we see that for the ε -insensitive distortion measure, the condition for $R_L^{(\varepsilon)}(D)$ to coincide with $R^{(\varepsilon)}(D)$ is the existence of a valid characteristic function $Q(\omega)$ satisfying

$$P(\omega) = Q(\omega) L_{|s|}(\omega) M_{|s|}^{(\varepsilon)}(\omega), \quad (17)$$

for the characteristic function $P(\omega)$ of the source distribution.

The above condition is rewritten as $P(\omega)/L_{|s|}(\omega) = M_{|s|}^{(\varepsilon)}(\omega)Q(\omega)$. If $Q(\omega)$ is the characteristic function of a probability distribution, that is, the Fourier transform of a density $q(y)$, then the right hand side is that of convolution of q and $m_{|s|}^{(\varepsilon)}$. Hence, Eq. (17) is never satisfied for any density $q(y)$ unless $P(\omega)/L_{|s|}(\omega)$ is the Fourier transform of a valid density. This means that a necessary condition for $R_L^{(\varepsilon)}(D_s) = R^{(\varepsilon)}(D_s)$ is given by $R^{(0)}(D_s) = R_L^{(0)}(D_s)$, that is, the SLB coincides with the rate-distortion function under the absolute-magnitude distortion ($\varepsilon = 0$).

We will use the above condition in Section III to prove $R_L^{(\varepsilon)}(D)$ is strictly smaller than $R^{(\varepsilon)}(D)$ for the Laplacian and Gaussian sources.

E. General Upper Bound

Let us turn to upper bounds for $R^{(\varepsilon)}(D)$. Since $\rho_\varepsilon(x) \leq \rho_0(x) = |x|$, we have a trivial upper bound,

$$R^{(\varepsilon)}(D) \leq R^{(0)}(D),$$

where $R^{(0)}(D)$ is the rate-distortion function for the absolute-magnitude distortion measure, $d(x, y) = |x - y|$.

Another more informative upper bound is obtained by taking $q(y|x) = g_s(y - x)$, where g_s is defined by Eq. (9), in the original rate-distortion problem in Eq. (1) [1, p. 103]. This yields the following upper bound,

$$R^{(\varepsilon)}(D_s) \leq R_U^{(\varepsilon)}(D_s) = h(r_s) - h(g_s), \quad (18)$$

where

$$r_s(y) = (g_s * p)(y) = \int g_s(y - x)p(x)dx \quad (19)$$

and D_s is given by Eq. (8) and further by Eq. (13). Note in Eq. (18) that the term $h(g_s)$ is common to the SLB and is specifically given by Eq. (12).

Since g_s is defined by ρ_ε as in Eq. (9), $h(r_s)$ is analytically intractable for many sources. Hence, we create a further upper bound which is analytically obtained for any sources with finite variance.

The Gaussian distribution with variance v maximizes differential entropy among the distributions whose variance is constrained to be v . The maximum value of differential entropy is $\frac{1}{2} \log(2\pi e v)$. Therefore, the differential entropy $h(r_s) = h(g_s * p)$ is upper bounded as follows,

$$h(r_s) \leq \frac{1}{2} \log \left(2\pi e (v_p + v_s^{(\varepsilon)}) \right), \quad (20)$$

where $v_p = \int x^2 p(x)dx - (\int x p(x)dx)^2$ and $v_s^{(\varepsilon)} = \int x^2 g_s(x)dx$. This is because the variance of $g_s * p$ is $v_p + v_s^{(\varepsilon)}$. The variance of g_s is specifically evaluated as,

$$v_s^{(\varepsilon)} = \frac{2}{C_s} \left\{ \frac{\varepsilon^3}{3} + \frac{1}{|s|} \left(\varepsilon^2 + \frac{2}{|s|}\varepsilon + \frac{2}{|s|^2} \right) \right\}, \quad (21)$$

where C_s is defined in Eq. (11). The general upper bound in Eq. (18), combined with Eqs. (20) and (21), yields the following upper bound to $R_U^{(\varepsilon)}(D)$,

$$R_{GE}^{(\varepsilon)}(D_s) = \frac{1}{2} \log \left(2\pi e (v_p + v_s^{(\varepsilon)}) \right) - h(g_s), \quad (22)$$

which is referred to as the Gaussian entropy bound.

In the next sections, we will evaluate these upper bounds for the Laplacian and Gaussian sources to examine the tightness of the general lower bound obtained in Theorem 1.

III. LAPLACIAN AND GAUSSIAN SOURCES

A. Laplacian Source

In this subsection, we consider the Laplacian source with parameter α ,

$$p(x) = l_\alpha(x) = \frac{\alpha}{2} e^{-\alpha|x|}. \quad (23)$$

The SLB for this source is given by Theorem 1 with the differential entropy,

$$h(p) = 1 - \log \frac{\alpha}{2}.$$

The maximum distortion in Eq. (5) is

$$D_{\max}^{(\varepsilon)} = \int \rho_\varepsilon(x)p(x)dx = \frac{1}{\alpha} e^{-\alpha\varepsilon}. \quad (24)$$

For the absolute-magnitude distortion measure ($\varepsilon = 0$),

$$R^{(0)}(D) = R_L^{(0)}(D) = -\log(\alpha D), \quad (0 \leq D \leq 1/\alpha), \quad (25)$$

holds [1, p. 95, Example 4.3.2.1] because the condition in Eq. (17) reduces to $M_{|s|}^{(0)}(\omega) = 1$ and

$$Q(\omega) = \frac{\alpha^2}{|s|^2} + \left(1 - \frac{\alpha^2}{|s|^2} \right) \frac{\alpha^2}{\alpha^2 + \omega^2},$$

which is the Fourier transform of the valid probability density, $q(y) = \frac{\alpha^2}{|s|^2} \delta(y) + \left(1 - \frac{\alpha^2}{|s|^2} \right) l_\alpha(y)$. For $\varepsilon > 0$, however, $R^{(\varepsilon)}(D)$ is strictly greater than $R_L^{(\varepsilon)}(D)$ for all D , which we will prove in the following. We will later derive an analytic upper bound to $R^{(\varepsilon)}(D)$ from Eq. (18).

The condition for $R^{(\varepsilon)}(D) = R_L^{(\varepsilon)}(D)$ in Eq. (17) is equivalent to

$$Q(\omega) = \frac{1}{M_{|s|}^{(\varepsilon)}(\omega)} \left\{ \frac{\alpha^2}{|s|^2} + \left(1 - \frac{\alpha^2}{|s|^2} \right) \frac{\alpha^2}{\alpha^2 + \omega^2} \right\}.$$

For $|s| > \alpha$, $\left\{ \frac{\alpha^2}{|s|^2} + \left(1 - \frac{\alpha^2}{|s|^2}\right) \frac{\alpha^2}{\alpha^2 + \omega^2} \right\} > \frac{\alpha^2}{|s|^2}$. Putting $\omega = (2k - 1/2)\frac{\pi}{\varepsilon}$, where k is a natural number, we have

$$\left| Q \left((2k - 1/2) \frac{\pi}{\varepsilon} \right) \right| > \frac{\alpha^2}{|s|^2} \frac{1 + \varepsilon|s|}{\varepsilon|s|} (2k - 1/2)\pi.$$

Since the right hand side becomes arbitrarily large for $k \rightarrow \infty$, $Q(\omega)$ can not be a Fourier transform of any density $q(y)$. This means that $R_L^{(\varepsilon)}(D_s) < R^{(\varepsilon)}(D_s)$ for $s < -\alpha$. It follows from Eq. (12) that $R_L^{(\varepsilon)}(D_{s=-\alpha}) = 1 - \log(1 + \alpha\varepsilon) - \frac{1}{1+\alpha\varepsilon} \leq 0$, which implies that $R^{(\varepsilon)}(D)$ is strictly greater than $R_L^{(\varepsilon)}(D)$ for all D .

To obtain an analytic upper bound for $R^{(\varepsilon)}(D)$, we evaluate $h(r_s)$ in Eq. (18). The density $r_s(y)$ defined in Eq. (19) is specifically given by

$$r_s(y) = \begin{cases} \frac{1}{2C_s} b_s(-y), & (y \leq -\varepsilon), \\ \frac{1}{2C_s} a_s(y), & (|y| < \varepsilon), \\ \frac{1}{2C_s} b_s(y), & (y \geq \varepsilon), \end{cases}$$

where $a_s(y) = \frac{s}{\alpha-s} e^{-\alpha(y+\varepsilon)} + \frac{s}{\alpha-s} e^{\alpha(y-\varepsilon)} + 2$ and $b_s(y) = \frac{s}{\alpha-s} e^{-\alpha(y+\varepsilon)} + \frac{s}{\alpha+s} e^{-\alpha(y-\varepsilon)} + \frac{2\alpha^2}{\alpha^2-s^2} e^{s(y-\varepsilon)}$. Since the differential entropy of $r_s(y)$ is not analytically simplified any more, we evaluate it from above to obtain an upper bound for $R^{(\varepsilon)}(D_s)$.

First, we bound $a_s(y)$ and $b_s(y)$ from below.

$$a_s(y) \geq 2 + \frac{2se^{-\alpha\varepsilon}}{\alpha-s} \cosh(\alpha\varepsilon) \equiv c_s$$

for $|y| < \varepsilon$ and since $s > -\alpha$,

$$b_s(y) \geq \frac{s}{\alpha-s} e^{-\alpha(y+\varepsilon)} + \frac{s}{\alpha+s} e^{-\alpha(y-\varepsilon)} + \frac{2\alpha^2}{\alpha^2-s^2} e^{-\alpha(y-\varepsilon)} \\ = c_s e^{\alpha(\varepsilon-y)}$$

for $y \geq \varepsilon$. Next, let us define

$$B_s \equiv \int_{\varepsilon}^{\infty} b_s(y) dy = \frac{s}{\alpha-s} \frac{1}{\alpha} e^{-2\alpha\varepsilon} + \frac{s^2(\alpha-s) - 2\alpha^3}{(\alpha^2-s^2)s\alpha}.$$

Finally, $h(r_s)$ is bounded as follows,

$$h(r_s) \leq -\log \frac{c_s}{2C_s} - \frac{\alpha\varepsilon}{C_s} B_s + \frac{\alpha}{C_s} E_s, \quad (26)$$

where

$$E_s \equiv \int_{\varepsilon}^{\infty} y b_s(y) dy \\ = \frac{s}{\alpha-s} \frac{1+\alpha\varepsilon}{\alpha^2} e^{-2\alpha\varepsilon} + \frac{s}{s+\alpha} \frac{1+\alpha\varepsilon}{\alpha^2} + \frac{2\alpha^2}{\alpha^2-s^2} \frac{1-s\varepsilon}{s^2},$$

which easily follows from $\int_{\varepsilon}^{\infty} y e^{-\alpha y} dy = \frac{1+\alpha\varepsilon}{\alpha^2} e^{-\alpha\varepsilon}$.

Thus, we have the further upper bound, $R_U^{(\varepsilon)}(D_s) \leq R_{AU}^{(\varepsilon)}(D_s)$, where

$$R_{AU}^{(\varepsilon)}(D_s) \equiv -\log \frac{c_s}{2C_s} - \frac{\alpha\varepsilon}{C_s} B_s + \frac{\alpha}{C_s} E_s - h(g_s), \quad (27)$$

which we will refer to as the analytic upper bound in Section IV.

In the low distortion limit, $D \rightarrow 0$ and $|s| \rightarrow \infty$, we have $C_s \rightarrow 2\varepsilon$, $c_s \rightarrow 2e^{-\alpha\varepsilon} \sinh(\alpha\varepsilon)$, $B_s \rightarrow \frac{2}{\alpha} e^{-\alpha\varepsilon} \sinh(\alpha\varepsilon)$, $E_s \rightarrow \frac{1+\alpha\varepsilon}{\alpha^2} 2e^{-\alpha\varepsilon} \sinh(\alpha\varepsilon)$. Then, Eq. (26) turns out to be

$$h(r_{-\infty}) \leq \alpha\varepsilon - \log \frac{\sinh(\alpha\varepsilon)}{2\varepsilon} + \frac{e^{-\alpha\varepsilon}}{\alpha\varepsilon} \sinh(\alpha\varepsilon).$$

Furthermore, it follows from $\frac{\sinh(\alpha\varepsilon)}{\alpha\varepsilon} = 1 + \frac{(\alpha\varepsilon)^2}{3!} + O(\varepsilon^4)$ and $e^{-\alpha\varepsilon} = 1 - \alpha\varepsilon + \frac{(\alpha\varepsilon)^2}{2} + O(\varepsilon^3)$ that

$$h(r_{-\infty}) \leq h(p) + \frac{(\alpha\varepsilon)^2}{2} + O(\varepsilon^3),$$

which implies that the SLB $R_L^{(\varepsilon)}(D)$ satisfies $R_L^{(\varepsilon)}(0) \leq R^{(\varepsilon)}(0) \leq R_L^{(\varepsilon)}(0) + \frac{(\alpha\varepsilon)^2}{2} + O(\varepsilon^3)$.

B. Gaussian Source

In this subsection, we consider the Gaussian source with mean zero and variance σ^2 ,

$$p(x) = \sqrt{\frac{1}{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}. \quad (28)$$

The differential entropy of this source is $h(p) = \frac{1}{2} \{1 + \log(2\pi\sigma^2)\}$. The maximum distortion in Eq. (5) is

$$D_{\max}^{(\varepsilon)} = \int \rho_{\varepsilon}(x) p(x) dx = 2 \{ \sigma^2 p(\varepsilon) - \varepsilon \Phi_c(\varepsilon/\sigma) \}, \quad (29)$$

where $\Phi_c(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt$. We show that $R^{(\varepsilon)}(D)$ for this source lies above its SLB for all D and consider small distortion limit of the Gaussian entropy upper bound in Eq. (22).

The condition for $R^{(\varepsilon)}(D_s) = R_L^{(\varepsilon)}(D_s)$ is given by Eq. (17) with $P(\omega) = e^{-\sigma^2\omega^2/2}$. The inverse transform of $P(\omega)/L_{|s|}(\omega) = \left(1 + \frac{\omega^2}{|s|^2}\right) P(\omega)$ is $p(x) \left(1 + \frac{1}{|s|^2\sigma^2} - \frac{x^2}{|s|^2\sigma^4}\right)$, which becomes negative for large x .² From the discussion below Eq. (17), this implies that $R^{(0)}(D) > R_L^{(0)}(D)$ and hence $R^{(\varepsilon)}(D) > R_L^{(\varepsilon)}(D)$ for all D .

We have the Gaussian entropy bound in Eq. (22) with $v_p = \sigma^2$. In the limit, $|s| \rightarrow \infty$, $v_s^{(\varepsilon)} \rightarrow \varepsilon^2/3$. Then, the upper bound in Eq. (20) is further upper bounded as follows,

$$\frac{1}{2} \log \left(2\pi e \left(\sigma^2 + \frac{\varepsilon^2}{3} \right) \right) \leq h(p) + \frac{\varepsilon^2}{6\sigma^2},$$

since $\log(1+x) \leq x$ for $x > 0$. This means that the SLB given by Theorem 1 provides an approximation to $R^{(\varepsilon)}(D)$ with accuracy $\frac{\varepsilon^2}{6\sigma^2}$ as $D \rightarrow 0$.

IV. NUMERICAL EVALUATION

Figure 2 depicts the functions $R_L^{(\varepsilon)}(D)$ and $R_{AU}^{(\varepsilon)}(D)$ for the Laplacian source in Eq. (23) with $\alpha = \sqrt{2}$ when $\varepsilon = 0.1$. It also shows $R_{GE}^{(\varepsilon)}(D)$ in Eq. (22) with $v_p = 2/\alpha^2 = 1$ and the trivial upper bound $R^{(0)}(D)$ given by Eq. (25). It is observed that the analytic upper bound and the SLB are very close to each other for small distortion ($D < 0.01$). The analytic upper

²See also Eq. (4.3.27) in [1, p. 97].

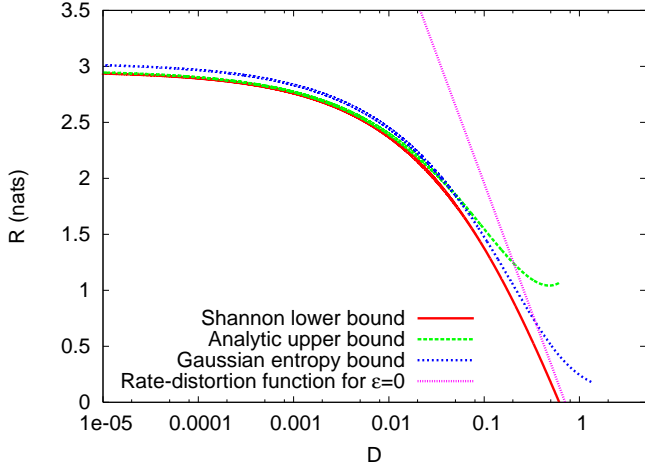


Fig. 2. Rate-distortion bounds for the Laplacian source.

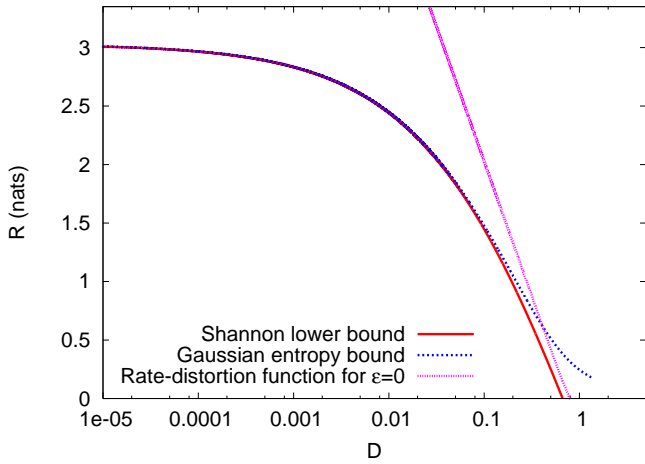


Fig. 3. Rate-distortion bounds for the Gaussian source.

bound becomes looser than the Gaussian entropy bound for large distortion ($0.05 < D$) while the trivial upper bound is relatively more informative about $R^{(\varepsilon)}(D)$ in the vicinity of $D_{\max}^{(\varepsilon)}$ when combined with the SLB. This suggests that the SLB provides reasonable approximation to $R^{(\varepsilon)}(D)$ even for large distortion. Let $D_{\max}^{(\varepsilon),L}$ denote the average distortion where the SLB $R_L^{(\varepsilon)}(D)$ reaches zero. We observed the following values of $D_{\max}^{(\varepsilon),L}$, $D_{\max}^{(\varepsilon)}$ defined in Eq. (24) and $D_{\max}^{(0)} = 1/\alpha$, $D_{\max}^{(\varepsilon),L} = 0.6136 < D_{\max}^{(\varepsilon)} = 0.6139 < D_{\max}^{(0)} = 0.7071$.

Figure 3 presents the bounds $R_L^{(\varepsilon)}(D)$ and $R_{GE}^{(\varepsilon)}(D)$ ($\varepsilon = 0.1$) and $R^{(0)}(D)$ for the Gaussian source in Eq. (28) with $\sigma^2 = 1$. The rate-distortion function $R^{(0)}(D)$ for the Gaussian source was evaluated according to its explicit parametric form given in [9, Th. 2]. It can be seen that the Gaussian entropy bound is very tight for small distortion ($D < 0.1$), which implies high accuracy of the SLB and that the trivial upper bound is informative about $R^{(\varepsilon)}(D)$ around $D = D_{\max}^{(\varepsilon)}$. The observed values of $D_{\max}^{(\varepsilon),L}$, $D_{\max}^{(\varepsilon)}$ defined in Eq. (29) and $D_{\max}^{(0)} = \sqrt{2\sigma^2/\pi}$ are as follows, $D_{\max}^{(\varepsilon),L} = 0.6662 < D_{\max}^{(\varepsilon)} = 0.7019 < D_{\max}^{(0)} = 0.7979$.

V. CONCLUSION

In this article, we have shown upper and lower bounds for the rate-distortion function of the ε -insensitive distortion measure. We derived the SLB, which is applicable to any source densities. Focusing on the Laplacian and Gaussian sources, we have proved that the rate-distortion functions for these sources are strictly greater than the corresponding SLBs for all D and provided upper bounds for the rate-distortion functions, which are proved to have accuracy of $O(\varepsilon^2)$ in the small distortion limit. We have demonstrated through numerical evaluation that the SLB is very accurate in the small distortion region while it still provides reasonable approximation to $R^{(\varepsilon)}(D)$ for the high distortion region around $D_{\max}^{(\varepsilon)}$ as $R^{(0)}(D)$ suggests.

In order to explicitly evaluate $R^{(\varepsilon)}(D)$, it would be important to explore properties of the optimal reproduction distribution $q_s(y)$. It is also important to develop practical learning algorithm for the mixture $\int g_s(x - y)q(y)dy$, where g_s is defined by the ε -insensitive loss function as in Eq. (10). Another issue to be addressed is the extension of our results to vector sources by using an extension of the ε -insensitive loss function to higher-dimensional vectors. There are some variations of the ε -insensitive loss function [3], [5]. It would also be an interesting undertaking to investigate properties of the rate-distortion functions for these variations.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 23700175.

REFERENCES

- [1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [2] A. Buzo, F. Kuhlmann and C. Rivera, "Rate-distortion bounds for quotient-based distortions with application to Itakura-Saito distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-32, no. 2, pp. 141–147, 1986.
- [3] W. Chu, S. S. Keerthi and C. J. Ong, "A unified loss function in Bayesian framework for support vector regression", *In Proc. 18th Int. Conf. Mach. Learn.*, pp. 51–58, 2001.
- [4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Interscience, 1991.
- [5] O. Dekel, S. Shalev-Shwartz and Y. Singer, " ε -insensitive regression by loss symmetrization," *J. Mach. Learn. Res.*, vol. 6, pp. 711–741, 2005.
- [6] R. M. Gray, *Entropy and Information Theory (2nd Ed.)*, Springer, 2011.
- [7] K. Rose, "A mapping approach to rate-distortion computation and analysis," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1939–1952, 1994.
- [8] I. Steinwart, A. Christmann, *Support Vector Machines*, Springer, 2008.
- [9] H. H. Tan and K. Yao, "Evaluation of rate-distortion functions for a class of independent identically distributed sources under an absolute magnitude criterion," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 1, pp. 59–64, 1975.
- [10] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.
- [11] K. Watanabe, S. Ikeda, "Rate-distortion function for gamma sources under absolute-log distortion," *In Proc. 35th Symp. Inform. Theory and its Appl.*, pp. 425–430, 2012.
- [12] K. Yao and H. H. Tan, "Absolute error rate-distortion functions for sources with constrained magnitudes," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 4, pp. 499–503, 1978.